Exam   :   **DP-200**

Title   :   Implementing an Azure Data Solution

Vendor   :   Microsoft

Version   :   DEMO

NO.1 You need to develop a pipeline for processing data. The pipeline must meet the following requirements.
* Scale up and down resources for cost reduction.
* Use an in-memory data processing engine to speed up ETL and machine learning operations.
* Use streaming capabilities.
* Provide the ability to code in SQL, Python, Scala, and R.
* Integrate workspace collaboration with Git.
What should you use?

A. HDInsight Spark Cluster

B. Azure SQL Data Warehouse

C. HDInsight Hadoop Cluster

D. Azure Stream Analytics

Answer: A

Explanation:
Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.
HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.
Languages: R, Python, Java, Scala, SQL
You can create an HDInsight Spark cluster using an Azure Resource Manager template. The template can be found in GitHub.
References:
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing

NO.2 Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You need to implement diagnostic logging for Data Warehouse monitoring.
Which log should you use?

A. DmsWorkers

B. SqlRequests

C. RequestSteps

D. ExecRequests

Answer: B

Explanation:
Scenario:
The Azure SQL Data Warehouse cache must be monitored when the database is being used.

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

References:
https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-pdw-sql-requests-transact-sq

NO.3 You are designing a new Lambda architecture on Microsoft Azure.
The real-time processing layer must meet the following requirements:
Ingestion:
* Receive millions of events per second
* Act as a fully managed Platform-as-a-Service (PaaS) solution
* Integrate with Azure Functions
Stream processing:
* Process on a per-job basis
* Provide seamless connectivity with Azure services
* Use a SQL-based query language
Analytical data store:
* Act as a managed service
* Use a document store
* Provide data encryption at rest
You need to identify the correct technologies to build the Lambda architecture using minimal effort.
Which technologies should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Architecture requirement   **Answer Area**

Ingestion

HDInsight Kafka	V
Azure Event Hubs	
HDInsight Storm	
HDInsight Spark	

Stream Processing

Azure Stream Analytics	V
HDInsight with Spark Streaming	
Azure Cosmos DB Change Feed	
Azure Analysis Services	

Analytical Data Store

Hive LLAP on HDInsight	V
Azure Analysis Services	
Azure Cosmos DB	
SQL Data Warehouse	

Answer:

Architecture requirement	Answer Area

Ingestion

HDInsight Kafka	V
Azure Event Hubs	
HDInsight Storm	
HDInsight Spark	

Stream Processing

Azure Stream Analytics	V
HDInsight with Spark Streaming	
Azure Cosmos DB Change Feed	
Azure Analysis Services	

Analytical Data Store

Hive LLAP on HDInsight	V
Azure Analysis Services	
Azure Cosmos DB	
SQL Data Warehouse	

Explanation:

Box 1: Azure Event Hubs

This portion of a streaming architecture is often referred to as stream buffering. Options include Azure Event Hubs, Azure IoT Hub, and Kafka.

Incorrect Answers: Not HDInsight Kafka

Azure Functions need a trigger defined in order to run. There is a limited set of supported trigger types, and Kafka is not one of them.

Box 2: Azure Stream Analytics

Azure Stream Analytics provides a managed stream processing service based on perpetually running SQL queries that operate on unbounded streams.

You can also use open source Apache streaming technologies like Storm and Spark Streaming in an HDInsight cluster.

Box 3: Azure SQL Data Warehouse

Azure SQL Data Warehouse provides a managed service for large-scale, cloud-based data warehousing. HDInsight supports Interactive Hive, HBase, and Spark SQL, which can also be used to serve data for analysis.

References:

https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/

NO.4 Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account. You need to toad the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create an external file format and external table using the external data source

3. Load the data using the INSERT...SELECT statement

Does the solution meet the goal?

A. No

B. Yes

Answer: A

Explanation:

You load the data using the CREATE TABLE AS SELECT statement.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

NO.5 A company runs Microsoft SQL Server in an on-premises virtual machine (VM).

You must migrate the database to Azure SQL Database. You synchronize users from Active Directory to Azure Active Directory (Azure AD).

You need to configure Azure SQL Database to use an Azure AD user as administrator. What should you configure?

A. For each Azure SQL Database, set the Access Control to administrator.

B. For the Azure SQL Database server, set the Active Directory to administrator.

C. For each Azure SQL Database, set the Active Directory administrator role.

D. For the Azure SQL Database server, set the Access Control to administrator.

Answer: C

Explanation:

There are two administrative accounts (Server admin and Active Directory admin) that act as administrators.

One Azure Active Directory account, either an individual or security group account, can also be configured as an administrator. It is optional to configure an Azure AD administrator, but an Azure AD administrator must be configured if you want to use Azure AD accounts to connect to SQL Database.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-manage-logins

NO.6 You develop data engineering solutions for a company.

A project requires the deployment of data to Azure Data Lake Storage.

You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Configure access control lists (ACL) for the Azure Data Lake Storage account.

B. Configure service-to-service authentication for the Azure Data Lake Storage account.

C. Assign Azure AD security groups to Azure Data Lake Storage.

D. Configure end-user authentication for the Azure Data Lake Storage account.

E. Create security groups in Azure Active Directory (Azure AD) and add project members.

Answer: A,C,E

Explanation:
References:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-secure-data

NO.7 You are developing a solution using a Lambda architecture on Microsoft Azure.
The data at test layer must meet the following requirements:
Data storage:
* Serve as a repository (or high volumes of large files in various formats.
* Implement optimized storage for big data analytics workloads.
* Ensure that data can be organized using a hierarchical structure.
Batch processing:
* Use a managed solution for in-memory computation processing.
* Natively support Scala, Python, and R programming languages.
* Provide the ability to resize and terminate the cluster automatically.
Analytical data store:
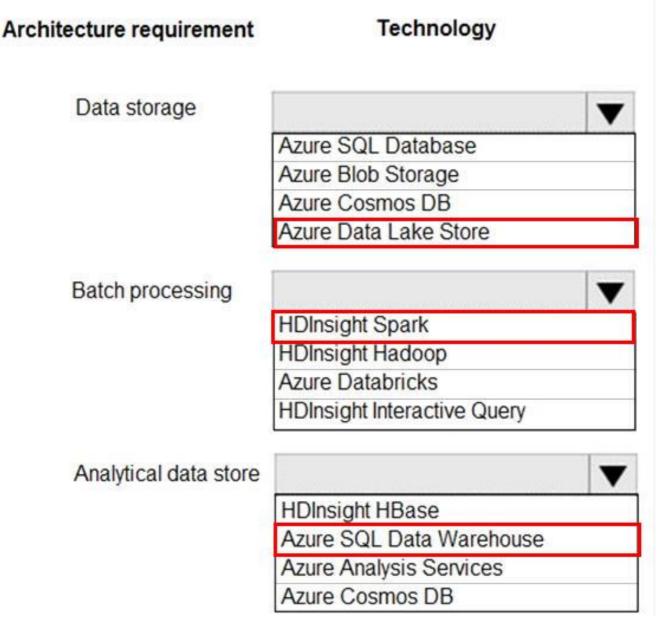* Support parallel processing.
* Use columnar storage.
* Support SQL-based languages.
You need to identify the correct technologies to build the Lambda architecture.
Which technologies should you use? To answer, select the appropriate options in the answer area
NOTE: Each correct selection is worth one point.

Architecture requirement	Technology
Data storage	▼
	Azure SQL Database
	Azure Blob Storage
	Azure Cosmos DB
	Azure Data Lake Store
Batch processing	▼
	HDInsight Spark
	HDInsight Hadoop
	Azure Databricks
	HDInsight Interactive Query
Analytical data store	▼
	HDInsight HBase
	Azure SQL Data Warehouse
	Azure Analysis Services
	Azure Cosmos DB

Answer:

Get Latest & Valid DP-200 Exam's Question and Answers from Realexamfree.com.

http://www.realexamfree.com/dp-200-real-exam-dumps.html

7

Architecture requirement	Technology
Data storage	▼
	Azure SQL Database
	Azure Blob Storage
	Azure Cosmos DB
	Azure Data Lake Store
Batch processing	▼
	HDInsight Spark
	HDInsight Hadoop
	Azure Databricks
	HDInsight Interactive Query
Analytical data store	▼
	HDInsight HBase
	Azure SQL Data Warehouse
	Azure Analysis Services
	Azure Cosmos DB

Explanation:
Data storage: Azure Data Lake Store
A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.
Batch processing: HD Insight Spark
Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.
HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.
Languages: R, Python, Java, Scala, SQL
Analytic data store: SQL Data Warehouse

SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).
SQL Data Warehouse stores data into relational tables with columnar storage.
References:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is

NO.8 You manage a solution that uses Azure HDInsight clusters.
You need to implement a solution to monitor cluster performance and status.
Which technology should you use?

A. Azure Log Analytics

B. Ambari REST API

C. Azure HDInsight REST API

D. Ambari Web UI

E. Azure HDInsight .NET SDK

Answer: D

Explanation:
Ambari is the recommended tool for monitoring utilization across the whole cluster. The Ambari dashboard shows easily glanceable widgets that display metrics such as CPU, network, YARN memory, and HDFS disk usage. The specific metrics shown depend on cluster type. The "Hosts" tab shows metrics for individual nodes so you can ensure the load on your cluster is evenly distributed.
The Apache Ambari project is aimed at making Hadoop management simpler by developing software for provisioning, managing, and monitoring Apache Hadoop clusters. Ambari provides an intuitive, easy-to-use Hadoop management web UI backed by its RESTful APIs.
References:
https://azure.microsoft.com/en-us/blog/monitoring-on-hdinsight-part-1-an-overview/
https://ambari.apache.org/